

Linear Precoding and Equalization for Network MIMO with Partial Cooperation

Saeed Kaviani, *Student Member, IEEE*, Osvaldo Simeone, *Senior Member, IEEE*, Witold A. Krzymieñ, *Senior Member, IEEE* and Shlomo Shamai (Shitz), *Fellow, IEEE*

arXiv:1302.2187v1 [cs.IT] 9 Feb 2013

Abstract—A cellular multiple-input multiple-output (MIMO) downlink system is studied in which each base station (BS) transmits to some of the users, so that each user receives its intended signal from a subset of the BSs. This scenario is referred to as network MIMO with partial cooperation, since only a subset of the BSs are able to coordinate their transmission towards any user. The focus of this paper is on the optimization of linear beamforming strategies at the BSs and at the users for network MIMO with partial cooperation. Individual power constraints at the BSs are enforced, along with constraints on the number of streams per user. It is first shown that the system is equivalent to a MIMO interference channel with generalized linear constraints (MIMO-IFC-GC). The problems of maximizing the sum-rate (SR) and minimizing the weighted sum mean square error (WSMSE) of the data estimates are non-convex, and suboptimal solutions with reasonable complexity need to be devised. Based on this, suboptimal techniques that aim at maximizing the sum-rate for the MIMO-IFC-GC are reviewed from recent literature and extended to the MIMO-IFC-GC where necessary. Novel designs that aim at minimizing the WSMSE are then proposed. Extensive numerical simulations are provided to compare the performance of the considered schemes for realistic cellular systems.

I. INTRODUCTION

Interference is known to be major obstacle for realizing the spectral efficiency increase promised by multiple-antenna techniques in wireless systems. Indeed, the multiple-input multiple-output (MIMO) capacity gains are severely degraded and limited in cellular environments due to the deleterious effect of interference [1],[2]. Therefore, network-level interference management appears to be of fundamental importance to overcome this limitation and harness the gains of MIMO technology. Confirming this point, multi-cell cooperation, also known as network MIMO, has been shown to significantly improve the system performance [3].

Manuscript received 24th June 2011; revised 14th November 2011; accepted 5th January 2012. The material in this paper was presented in part at the IEEE Global Communication Conference (GLOBECOM'11), Houston, U.S.A., Dec. 2011.

S. Kaviani and W. A. Krzymieñ are with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada (e-mails: skaviani@ualberta.ca,wak@ece.ualberta.ca). They are also with TRLabs, Edmonton, AB, Canada. The work of S. Kaviani and W. A. Krzymieñ was supported by TRLabs, Rohit Sharma Professorship, and the Natural Sciences and Engineering Research Council (NSERC) of Canada.

O. Simeone is with the CWCSPR, New Jersey Institute of Technology, Newark, NJ, USA (e-mail: osvaldo.simeone@njit.edu). The work of O. Simeone was supported by the U.S. N.S.F. under grant CCF-0914899.

S. Shamai (Shitz) is with the Department of Electrical Engineering, Technion-Israel Institute of Technology, Haifa, Israel (e-mail: sshlomo@ee.technion.ac.il). The work of S. Shamai was supported by the Israel Science Foundation (ISF).

Network MIMO involves cooperative transmission by multiple base stations (BSs) to each user. Depending on the level of multi-cell cooperation, network MIMO reduces to a number of scenarios, ranging from a MIMO broadcast channel (BC) [4] in case of full cooperation among all BSs, to a MIMO interference channel [5],[6] in case no cooperation at the BSs is allowed. In general, network MIMO allows cooperation only among a cluster of BSs for transmission to a certain user [7], [8] (see also references in [3]).

In this paper, we consider a MIMO interference channel with partial cooperation at the BSs. It is noted that all BSs cooperating for transmission to a certain user have to be informed about the message (i.e., the bit string) intended for the user. This can be realized using the backhaul links among the BSs and the central switching unit. We focus on the sum-rate maximization (SRM) and on the minimization of weighted sum-MSE (WSMSE) under per-BS power constraints and constraints on the number of streams per user. Moreover, although non-linear processing techniques such as vector precoding [9], [10] may generally be useful, we focus on more practical linear processing techniques. Both the SRM and WSMSE minimization (WSMMSE) problems are non-convex [11], and thus suboptimal design strategies of reasonable complexity are called for.

The contributions of this paper are as follows:

- (i) It is first shown in Sec. II-A that network MIMO with partial BS cooperation, that is, with partial message knowledge, is equivalent to a MIMO interference channel in which each transmitter knows the message of only one user under generalized linear constraints, which we refer to as MIMO-IFC-GC;
- (ii) We review the available suboptimal techniques that have been proposed for the SRM problem [12]–[14] and extend them to the MIMO-IFC-GC scenario where necessary in Sec. V. Since these techniques are generally unable to enforce constraints on the number of streams, we also review and generalize techniques that are based on the idea of interference alignment [5] and are able to impose such constraints;
- (iii) Then, we propose two novel suboptimal solutions for the WSMMSE problem in Sec. VI under arbitrary constraints on the number of streams. It is noted that the WSMMSE problem without such constraints would be trivial, as it would result in zero MMSE and no stream transmitted. The proposed solutions are based on a novel insight into the single-user MMSE problem with multiple linear constraints, which is discussed in Sec. IV;

- (iv) Finally, extensive numerical simulations are provided in Sec. VII to compare performance of the proposed schemes in realistic cellular systems.

Linear MMSE precoding and equalization techniques proposed in this paper were discussed briefly in [15]. The detailed analysis and discussion (including the proofs to the lemmas) are included in this paper. Additionally, we have also reviewed and extended available solutions to the SRM problem. Furthermore, we have included discussions of the complexity and overhead of the proposed techniques and previously available (and/or extended) solutions.

Notation: We denote the positive definite matrices as $\mathbf{A} \succeq 0$. $[\cdot]^+$ denotes $\max(\cdot, 0)$. Capital bold letters represent matrices and small bold letters represent vectors. We denote the transpose operator with $(\cdot)^T$ and conjugate transpose (Hermitian) with $(\cdot)^H$. $\mathbf{A}^{-\frac{1}{2}}$ represents the inverse square of positive definite matrix \mathbf{A} .

II. SYSTEM MODEL AND PRELIMINARIES

Consider the MIMO downlink system illustrated in Fig. 1 with M base stations (BSs) forming a set \mathcal{M} , and K users forming a set \mathcal{K} . Each BS is equipped with n_t transmit antennas and each mobile user employs n_r receive antennas. The m th BS is provided with the messages of its assigned users set $\mathcal{K}_m \subseteq \mathcal{K}$. In other words, the k th user receives its message from a subset of M_k BSs $\mathcal{M}_k \subseteq \mathcal{M}$. Notice that, if \mathcal{K}_m contains one user for each transmitter m and $M_k = 1$, then the model at hand reduces to a standard MIMO interference channel. Moreover, when all transmitters cooperate in transmitting to all the users, i.e., $\mathcal{K}_m = \mathcal{K}$ for all $m \in \mathcal{M}$ or equivalently $M_k = M$, then we have a MIMO broadcast channel (BC).

We now detail the signal model for the channel at hand, which is referred to as *MIMO interference channel with partial message sharing*. Define as $\mathbf{u}_k = [u_{k,1} \dots u_{k,d_k}]^T \in \mathbb{C}^{d_k}$ the $d_k \times 1$ complex vector representing the $d_k \leq \min(M_k n_t, n_r)$ independent information streams intended for user k . We assume that $\mathbf{u}_k \sim \mathcal{CN}(0, \mathbf{I})$. The data streams \mathbf{u}_k are known to all the BS in the set \mathcal{M}_k . In particular, if $m \in \mathcal{M}_k$, the m th

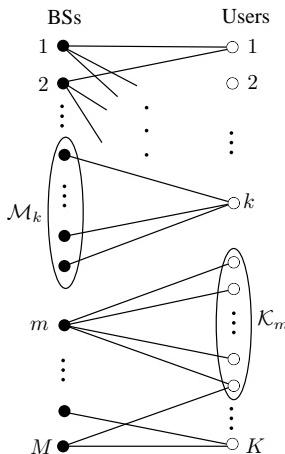


Fig. 1. A downlink model with partial BS cooperation or equivalently partial message knowledge.

BS precodes vector \mathbf{u}_k via a matrix $\mathbf{B}_{k,m} \in \mathbb{C}^{n_t \times d_k}$, so that the signal $\tilde{\mathbf{x}}_m \in \mathbb{C}^{n_t}$ sent by the m th BS can be expressed as

$$\tilde{\mathbf{x}}_m = \sum_{k \in \mathcal{K}_m} \mathbf{B}_{k,m} \mathbf{u}_k. \quad (1)$$

Imposing a per-BS power constraint, the following constraint must be then satisfied

$$\begin{aligned} \mathbb{E} [| | \tilde{\mathbf{x}}_m | |^2] &= \text{tr} \{ \mathbb{E} [\tilde{\mathbf{x}}_m \tilde{\mathbf{x}}_m^H] \} \\ &= \sum_{k \in \mathcal{K}_m} \text{tr} \{ \mathbf{B}_{k,m} \mathbf{B}_{k,m}^H \} \leq P_m, m = 1, \dots, M, \end{aligned} \quad (2)$$

where P_m is the power constraint of the m th BS.

The signal received at the k th user can be written as

$$\mathbf{y}_k = \sum_{m=1}^M \tilde{\mathbf{H}}_{k,m} \tilde{\mathbf{x}}_m + \tilde{\mathbf{n}}_k \quad (3a)$$

$$= \sum_{m \in \mathcal{M}_k} \tilde{\mathbf{H}}_{k,m} \mathbf{B}_{k,m} \mathbf{u}_k + \sum_{l \neq k} \sum_{j \in \mathcal{M}_l} \tilde{\mathbf{H}}_{k,j} \mathbf{B}_{l,j} \mathbf{u}_l + \tilde{\mathbf{n}}_k, \quad (3b)$$

where $\tilde{\mathbf{H}}_{k,m} \in \mathbb{C}^{n_r \times n_t}$ is the channel matrix between the m th BS and k th user and $\tilde{\mathbf{n}}_k$ is additive complex Gaussian noise $\tilde{\mathbf{n}}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ ¹. We assume ideal channel state information at all nodes. In (3b), we have distinguished between the first term, which represents useful signal, the second term, which accounts for interference, and the noise.

A. Equivalence with MIMO-IFC-GC

We now show that the MIMO interference channel with *partial message sharing and per-BS power constraints* described above is equivalent to a specific MIMO interference channel with *individual message knowledge and generalized linear constraints*, which we refer to as MIMO-IFC-GC.

Definition 1: (MIMO-IFC-GC) The MIMO-IFC-GC consists of K transmitters and K receiver, where the k th transmitter has $m_{t,k}$ antennas and the k th receiver has $m_{r,k}$ antennas. The received signal at the k th receiver is

$$\mathbf{y}_k = \mathbf{H}_{k,k} \mathbf{x}_k + \sum_{l \neq k} \mathbf{H}_{k,l} \mathbf{x}_l + \mathbf{n}_k, \quad (4)$$

where $\mathbf{n}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$, the inputs are $\mathbf{x}_k \in \mathbb{C}^{m_{t,k}}$ and the channel matrix between the l th transmitter and the k th receiver is $\mathbf{H}_{k,l} \in \mathbb{C}^{m_{r,k} \times m_{t,k}}$. The data vector intended for user k is $\mathbf{u}_k \in \mathbb{C}^{d_k}$ with $d_k \leq \min(m_{t,k}, m_{r,k})$ and $\mathbf{u}_k \sim \mathcal{CN}(0, \mathbf{I})$. The precoding matrix for user k is defined as $\mathbf{B}_k \in \mathbb{C}^{m_{t,k} \times d_k}$ so that $\mathbf{x}_k = \mathbf{B}_k \mathbf{u}_k$. The inputs \mathbf{x}_k have to satisfy M generalized linear constraints

$$\sum_{k=1}^K \text{tr} \{ \Phi_{k,m} \mathbb{E} [\mathbf{x}_k \mathbf{x}_k^H] \} = \sum_{k=1}^K \text{tr} \{ \Phi_{k,m} \mathbf{B}_k \mathbf{B}_k^H \} \leq P_m, \quad (5)$$

for given weight matrices $\Phi_{k,m} \in \mathbb{C}^{m_{t,k} \times m_{t,k}}$ and $m = 1, \dots, M$. The weight matrices are such that matrices $\sum_{m=1}^M \Phi_{k,m}$ are positive definite for all $k = 1, \dots, K$.

¹In case the noise is not uncorrelated across the antennas, each user can always whiten it as a linear pre-processing step. Therefore, a spatially uncorrelated noise can be assumed without loss of generality.

We remark that the positive definiteness of matrices $\sum_{m=1}^M \Phi_{k,m}$ guarantees that the system is not allowed to transmit infinite power in any direction [16].

Lemma 1: Let $(l)_k$ be the l th BS in subset \mathcal{M}_k of BSs that know user k 's message. The MIMO interference channel with partial message sharing (and per-transmitter power constraints) is equivalent to a MIMO-IFC-GC. This equivalent MIMO-IFC-GC is defined with $m_{t,k} = M_k n_t$, $m_{r,k} = n_r$, channel matrices

$$\mathbf{H}_{k,l} = \left[\tilde{\mathbf{H}}_{k,(1)_l} \cdots \tilde{\mathbf{H}}_{k,(M_l)_l} \right], \quad (6)$$

beamforming matrices

$$\mathbf{B}_k = \left[\mathbf{B}_{k,(1)_k}^\top \cdots \mathbf{B}_{k,(M_k)_k}^\top \right]^\top \quad (7)$$

and weight matrices $\Phi_{k,m}$ being all zero except that their l th $n_t \times n_t$ submatrix on the main diagonal is \mathbf{I}_{n_t} , if $m = (l)_k$ (If $k \notin \mathcal{K}_m$ then $\Phi_{k,m} = \mathbf{0}$). We emphasize that the definition of MIMO-IFC-GC and this equivalence rely on the assumption of linear processing at the transmitters.

Proof: : The proof follows by inspection. Notice that matrices $\sum_{m=1}^M \Phi_{k,m}$ are positive definite by construction. \square

Given the generality of the MIMO-IFC-GC, which includes the scenario of interest of MIMO interference channel with partial message sharing as per the Lemma above, in the following we focus on the MIMO-IFC-GC as defined above and return to the cellular application in Sec. VII. It is noted that a model that subsumes the MIMO-IFC-GC has been studied in [16], as discussed below.

B. Linear Receivers and Mean Square Error

In this paper, we focus on the performance of the MIMO-IFC-GC under linear processing at the receivers. Therefore, the k th receiver estimates the intended vector \mathbf{u}_k using the receive processing (or equalization) matrix $\mathbf{A}_k \in \mathbb{C}^{d_k \times m_{r,k}}$ as

$$\hat{\mathbf{u}}_k = \mathbf{A}_k^\top \mathbf{y}_k. \quad (8)$$

The most common performance measures, such as weighted sum-rate or bit error rate, can be derived from the estimation error covariance matrix for each user k ,

$$\mathbf{E}_k = \mathbb{E} \left[(\hat{\mathbf{u}}_k - \mathbf{u}_k) (\hat{\mathbf{u}}_k - \mathbf{u}_k)^\top \right], \quad (9)$$

which is referred to as *Mean Square Error (MSE)-matrix* (see [17] for a review). The name comes from the fact that the j th term on the main diagonal of \mathbf{E}_k is the MSE

$$\text{MSE}_{k,j} = \mathbb{E}[|\hat{u}_{k,j} - u_{k,j}|^2] \quad (10)$$

on the estimation of the k th user's j th data stream $u_{k,j}$. Using the definition of MIMO-IFC-GC, it is easy to see that the MSE-matrix can be written as a function of the equalization matrix \mathbf{A}_k and *all* the transmit matrices $\{\mathbf{B}_k\}_{k=1}^K$ as

$$\begin{aligned} \mathbf{E}_k = & \mathbf{A}_k^\top \mathbf{H}_k \mathbf{B}_k \mathbf{B}_k^\top \mathbf{H}_{k,k}^\top \mathbf{A}_k - \mathbf{A}_k^\top \mathbf{H}_{k,k} \mathbf{B}_k \\ & - \mathbf{B}_k^\top \mathbf{H}_{k,k}^\top \mathbf{A}_k + \mathbf{A}_k^\top \boldsymbol{\Omega}_k \mathbf{A}_k + \mathbf{I}_k. \end{aligned} \quad (11)$$

where $\boldsymbol{\Omega}_k$ is the covariance matrix that accounts for noise and interference at user k

$$\boldsymbol{\Omega}_k = \mathbf{I} + \sum_{l \neq k} \mathbf{H}_{k,l} \mathbf{B}_l \mathbf{B}_l^\top \mathbf{H}_{k,l}^\top. \quad (12)$$

III. PROBLEM DEFINITION AND PRELIMINARIES

In this paper, we consider the optimization of the sum of some specific functions $f_k(\mathbf{E}_k)$ of the MSE-matrices \mathbf{E}_k of all users $k = 1, \dots, K$ for the MIMO-IFC-GC. Specifically, we address the following constrained optimization problem

$$\begin{aligned} & \text{minimize}_{\mathbf{B}_k, \mathbf{A}_k, \forall k} \sum_{k=1}^K f_k(\mathbf{E}_k) \\ & \text{subject to} \quad \sum_{k=1}^K \text{tr} \{ \Phi_{k,m} \mathbf{B}_k \mathbf{B}_k^\top \} \leq P_m, m = 1, \dots, M, \end{aligned} \quad (13)$$

where the optimization is over all transmit beamforming matrices \mathbf{B}_k and equalization matrices \mathbf{A}_k . Specifically, we focus on the *weighted sum-MSE functions* (WSMSE)

$$f_k(\mathbf{E}_k) = \text{tr} \{ \mathbf{W}_k \mathbf{E}_k \} = \sum_{j=1}^{d_k} w_{kj} \text{MSE}_{kj} \quad (14)$$

with given diagonal weight matrices $\mathbf{W}_k \in \mathbb{C}^{d_k \times d_k}$ where the main diagonal of \mathbf{W}_k is given by $[w_{k,1}, \dots, w_{k,d_k}]$ with non-negative weights $w_{kj} \geq 0$. With cost function (14), we refer to the problem (13) as the *weighted sum-MSE minimization (WSMMSE)* problem.

Of more direct interest for communications systems is the maximization of the sum-rate. This is obtained from (13) by selecting the *sum-rate (SR) functions*

$$f_k(\mathbf{E}_k) = \log |\mathbf{E}_k|. \quad (15)$$

With cost function (15), problem (13) is referred to as the *sum-rate maximization (SRM)* problem. In fact, from information-theoretic considerations, it can be seen that (15) is the maximum achievable rate (in bits per channel use) for the k th user where the signals of the other users are treated as noise (see, e.g., [17]).

Remark 1: Consider an iterative algorithm where at each iteration a WSMMSE problem is solved with the weight matrices \mathbf{W}_k assumed to be non-diagonal and selected based on the previous MSE-matrix \mathbf{E}_k . This algorithm can approximate the solution of (13) for any general cost function $f_k(\mathbf{E}_k)$. This was first pointed out in [18] for the weighted SRM problem in a MIMO BC, then in [19] for the single-antenna interference channel and a general utility function, and has been generalized to a MIMO (broadcast) interference channel in [20] with conventional power constraints. It is not difficult to see that this result extends also to the MIMO-IFC-GC, which is not subsumed in the model of [20] due to the generalized linear constraints. We explicitly state this conclusion below.

Lemma 2 [20]: For strictly concave utility functions $f_k(\cdot)$ for all k , the global optimal solution of problem (13) and the solution of

$$\begin{aligned} & \text{minimize}_{\mathbf{B}_k, \mathbf{A}_k, \mathbf{W}_k, \forall k} \sum_{k=1}^K \{ \text{tr} \{ \mathbf{W}_k \mathbf{E}_k \} - \text{tr} \{ \mathbf{W}_k g_k(\mathbf{W}_k) \} \\ & \quad + f_k(g_k(\mathbf{W}_k)) \} \\ & \text{subject to} \quad \sum_{k=1}^K \text{tr} \{ \Phi_{k,m} \mathbf{B}_k \mathbf{B}_k^\top \} \leq P_m, m = 1, \dots, M, \end{aligned} \quad (16)$$

where $g_k(\cdot)$ is the inverse function of the $\nabla f_k(\cdot)$, are the same.

Consequently, in order to find an approximate solution of (13), at each step matrices \mathbf{W}_k for $k = 1, \dots, K$ are updated by solving (16) with respect to \mathbf{W}_k only (i.e., we keep $(\mathbf{A}_k, \mathbf{B}_k), \forall k$ unchanged in this step). Then, using the obtained matrices \mathbf{W}_k , for $k = 1, \dots, K$, the problem (16) reduces to a WSMMSE problem with respect to matrices \mathbf{A}_k and \mathbf{B}_k for $k = 1, \dots, K$ (i.e., matrices \mathbf{W}_k are kept fixed). This results in the iterative algorithm, that is discussed in Remark 1 and that leads to a suboptimal solution of (13). In the special case of the SRM problem, we have $f_k(\mathbf{E}_k) = \log |\mathbf{E}_k|$ and $g_k(\mathbf{W}_k) = \mathbf{W}_k^{-1}$, in which problem (16) is then equivalent to the problem

$$\begin{aligned} & \underset{\mathbf{B}_k, \mathbf{A}_k, \mathbf{W}_k, \forall k}{\text{minimize}} \quad \sum_{k=1}^K \text{tr} \{ \mathbf{W}_k \mathbf{E}_k \} - \sum_{k=1}^K \log |\mathbf{W}_k| \\ & \text{subject to} \quad \sum_{k=1}^K \text{tr} \{ \Phi_{k,m} \mathbf{B}_k \mathbf{B}_k^H \} \leq P_m, m = 1, \dots, M. \end{aligned} \quad (17)$$

The optimization problem (17) can be solved in an iterative fashion, where at each iteration the weights are selected as $\mathbf{W}_k^* = \mathbf{E}_k^{-1}$ and then the WSMMSE problem is solved with respect to matrices $(\mathbf{A}_k, \mathbf{B}_k)$ for $k = 1, \dots, K$.

IV. THE SINGLE-USER CASE ($K = 1$)

The WSMMSE and SRM problems are non-convex and thus global optimization is generally prohibitive. In this section, we address the case of a single user ($K = 1$). In particular, the SRM problem with $K = 1$ is non-convex if one includes constraints on the number of streams d_1 , but is otherwise convex and in this special case can be solved efficiently [17]. The global optimal solution for the single-user problem with multiple linear power constraint (and a rank constraint) is still unknown [21]. The WSMMSE problem is trivial without rank constraint, as explained above, and is non-convex. Here we first review a key result in [17][22] that shows with $K = 1$ and a single constraint ($M = 1$) the solution of the WSMMSE problem can be, however, found efficiently. We then discuss that with multiple constraints ($M > 1$), this is not the case, and a solution of the WSMMSE problem even with $K = 1$ must be found through some complex global optimization strategies. One such technique was recently proposed in [21] based on a sophisticated gradient approach. At the end of this section we then propose a computationally and conceptually simpler solution based on a novel result (Lemma 5), that our numerical result have shown to have excellent performance. This will be then leveraged in Sec. VI-B to propose a novel solution for the general multiuser case.

To elaborate, consider a scenario where the noise-plus-interference matrix Ω_k (12) is fixed and given (i.e., not subject to optimization). Now, we solve the WSMMSE problem with $K = 1$ for specified weight matrices \mathbf{W} and Φ_m . For the rest of this section, we drop the index $k = 1$ from all quantities for simplicity of notation. We proceed by solving the problem at hand, first with respect to \mathbf{A} for fixed \mathbf{B} , and then with respect to \mathbf{B} without loss of optimality. The first optimization, over \mathbf{A} , is easily seen to be a convex problem (without constraints) whose solution is given by the minimum MSE equalization

matrix

$$\mathbf{A} = (\mathbf{H} \mathbf{B} \mathbf{B}^H \mathbf{H}^H + \Omega)^{-1} \mathbf{H} \mathbf{B}. \quad (18)$$

Plugging (18) in the MSE matrix (11), we obtain

$$\mathbf{E} = \left(\mathbf{I} + \mathbf{B}^H \mathbf{H}^H \Omega^{-1} \mathbf{H} \mathbf{B} \right)^{-1}. \quad (19)$$

We now need to optimize over \mathbf{B} the following problem

$$\begin{aligned} & \underset{\mathbf{B}}{\text{minimize}} \quad \text{tr} \left\{ \mathbf{W} \left(\mathbf{I} + \mathbf{B}^H \mathbf{H}^H \Omega^{-1} \mathbf{H} \mathbf{B} \right)^{-1} \right\} \\ & \text{subject to} \quad \text{tr} \left\{ \Phi_m \mathbf{B} \mathbf{B}^H \right\} \leq P_m, m = 1, \dots, M \end{aligned}, \quad (20)$$

Consider first the single-constraint problem, i.e., $M = 1$. The global optimal solution for single-user WSMMSE problem with $M = 1$ is given in [22][21] and reported below. Recall that, according to Definition 1, matrix Φ_1 is positive definite.

Lemma 3 [22]: The optimal solution of the WSMMSE problem with $K = 1$ and a single trace constraint ($M = 1$) is given by

$$\mathbf{B} = \Phi_1^{-\frac{1}{2}} \mathbf{U} \Sigma, \quad (21)$$

where $\mathbf{U} \in \mathbb{C}^{m \times d}$ is the matrix of eigenvectors of matrix $\Phi_1^{-\frac{1}{2}} \mathbf{H} \Omega^{-1} \mathbf{H}^H \Phi_1^{-\frac{1}{2}}$ corresponding to its largest eigenvalues $\gamma_1 \geq \dots \geq \gamma_d$ and Σ is a diagonal matrix with the diagonal terms $\sqrt{p_i}$ defined as

$$p_i = \left[\sqrt{\frac{w_i}{\mu \gamma_i}} - \frac{1}{\gamma_i} \right]^+, \quad (22)$$

with $\mu \geq 0$ being the “waterfilling” level chosen so as to satisfy the single power constraint $\text{tr} \{ \Phi_1 \mathbf{B} \mathbf{B}^H \} = P_1$.

Proof: Introducing the “effective” precoding matrix $\bar{\mathbf{B}} = \Phi_1^{1/2} \mathbf{B}$ and “effective” channel matrix $\bar{\mathbf{H}} = \mathbf{H} \Phi_1^{-\frac{1}{2}}$, the problem is equivalent to the one discussed in [22, Theorem 1]. \square

In the case of multiple constraints the approach used in Lemma 3 cannot be leveraged. Here we propose a simple, but effective, approach, which is based on the following considerations summarized in the following two lemmas.

Lemma 4: The precoding matrix (21)-(22) for a given fixed $\mu > 0$ minimizes the Lagrangian function

$$\begin{aligned} \mathcal{L}(\bar{\mathbf{B}}; \mu) = & \text{tr} \left\{ \mathbf{W} \left(\mathbf{I} + \bar{\mathbf{B}}^H \Phi_1^{-\frac{1}{2}} \mathbf{H}^H \Omega^{-1} \mathbf{H} \Phi_1^{-\frac{1}{2}} \bar{\mathbf{B}} \right)^{-1} \right\} \\ & + \mu \text{tr} \{ \bar{\mathbf{B}} \bar{\mathbf{B}}^H \} \end{aligned} \quad (23)$$

where $\bar{\mathbf{B}}$ is the effective precoding matrix defined above.

Proof: We first note that (23) is the Lagrangian function of the single-user single-constraint problem solved in Lemma 2. Then, we prove (23) by contradiction. Assume that the minimum of the Lagrangian function is attained at where the corresponding \mathbf{E} is not diagonal. Then, one can always find a unitary matrix $\mathbf{Q} \in \mathbb{C}^{d \times d}$ such that the matrix $\bar{\mathbf{B}}^* = \bar{\mathbf{B}} \mathbf{Q}$ diagonalizes \mathbf{E} since with $\bar{\mathbf{B}}^*$ we have $\mathbf{E} = \mathbf{Q}^H \left(\mathbf{I} + \bar{\mathbf{B}}^H \Phi_1^{-\frac{1}{2}} \mathbf{H}^H \Omega^{-1} \mathbf{H} \Phi_1^{-\frac{1}{2}} \bar{\mathbf{B}} \right)^{-1} \mathbf{Q}$ [22]. The function $\text{tr} \{ \mathbf{W} \mathbf{E} \}$ is Schur concave, and therefore the matrix $\bar{\mathbf{B}}^*$ does not decrease the function $\text{tr} \{ \mathbf{W} \mathbf{E} \}$ with respect to $\bar{\mathbf{B}}$, while $\bar{\mathbf{B}} \bar{\mathbf{B}}^H = \bar{\mathbf{B}}^* \bar{\mathbf{B}}^{*H}$. This implies that the minimum of $\text{tr} \{ \mathbf{W} \mathbf{E} \}$

is reached when the MSE matrix is diagonalized. Therefore, we can set without loss of generality $\tilde{\mathbf{B}} = \mathbf{U}\Sigma$ where \mathbf{U} is defined as in Lemma 3 and Σ is diagonal with non-negative elements on the main diagonal. Substituting this form of $\tilde{\mathbf{B}}$ into the Lagrangian function, we obtain a convex problem in the diagonal elements of Σ , whose solution can be easily shown to be given by (22) for the given μ . This concludes the proof. \square

Lemma 5: Let p^* be the optimal value of the single-user WSMMSE problem with multiple constraints ($K = 1, M \geq 1$). We have

$$p^* \geq \max_{\lambda \geq 0} \inf_{\mathbf{B}} \mathcal{L}(\mathbf{B}; \lambda), \quad (24)$$

where

$$\begin{aligned} \mathcal{L}(\mathbf{B}; \lambda) &= \text{tr} \left\{ \mathbf{W} (\mathbf{I} + \mathbf{B}^H \mathbf{H}^H \boldsymbol{\Omega} \mathbf{H} \mathbf{B})^{-1} \right\} \\ &\quad + \sum_{m=1}^M \lambda_m (\text{tr} \{ \boldsymbol{\Phi}_m \mathbf{B} \mathbf{B}^H \} - P_m) \end{aligned} \quad (25)$$

is the Lagrangian function of the single-user WSMMSE problem at hand and $\lambda = (\lambda_1, \dots, \lambda_M)$. Moreover, if there exists an optimal solution $\tilde{\mathbf{B}}$ achieving p^* that, together with a strictly positive Lagrange multiplier $\tilde{\lambda} > 0$, satisfies the conditions

$$\nabla_{\mathbf{B}} \mathcal{L} = 0, \quad (26)$$

$$\text{tr} \{ \boldsymbol{\Phi}_m \tilde{\mathbf{B}} \tilde{\mathbf{B}}^H \} = P_m, \quad \forall m \quad (27)$$

then (24) holds with equality.

Proof: The proof is given in the appendix. \square

Lemma 5 suggests that to solve the single-user multiple-constraint problem, under some technical conditions, one can minimize instead the dual problem on the right-hand side of (24). Lemma 3 showed that this is always possible with a single constraint. The conditions in Lemma 5 hold in most cases where the power constraints for the optimal solution are satisfied with equality. While this may not be always the case, in practice, e.g., if the power constraints represent per-BS power constraints as in the original formulation of Sec. II, this condition can be shown to hold [23].

Inspired by Lemma 5, here we propose an iterative approach to the solution of the WSMMSE problem with $K = 1$ that is based on solving the dual problem $\max_{\lambda \geq 0} \min_{\mathbf{B}} \mathcal{L}(\mathbf{B}; \lambda)$. Specifically, in order to maximize $\inf_{\mathbf{B}} \mathcal{L}(\mathbf{B}; \lambda)$ over $\lambda \succeq 0$, in the proposed algorithm, the auxiliary variables λ is updated at the j th iteration via a subgradient update given by [16]

$$\lambda_m^{(j)} = \lambda_m^{(j-1)} + \delta (P_m - \text{tr} \{ \boldsymbol{\Phi}_m \mathbf{B} \mathbf{B}^H \}), \quad \forall m, \quad (28)$$

so as to attempt to satisfy the power constraints. Having fixed the vector $\lambda^{(j)}$, problem $\min_{\mathbf{B}} \mathcal{L}(\mathbf{B}, \lambda)$ reduces to minimizing (23) with $\boldsymbol{\Phi}_1 = \boldsymbol{\Phi}(\lambda^{(j)}) = \sum_m \lambda_m^{(j)} \boldsymbol{\Phi}_m$ and $\mu = 1$. This can be done using Lemma 3, so that from (21)-(22), at the j th iteration, $\mathbf{B}^{(j)}$ is obtained as $\boldsymbol{\Phi}(\lambda^{(j)})^{-\frac{1}{2}} \mathbf{U}^{(j)} \boldsymbol{\Sigma}^{(j)}$ where $\mathbf{U}^{(j)}$ is the matrix of eigenvectors of matrix $\boldsymbol{\Phi}(\lambda^{(j)})^{-\frac{1}{2}} \mathbf{H} \mathbf{H}^H \boldsymbol{\Phi}(\lambda^{(j)})^{-\frac{1}{2}}$ corresponding to its largest eigenvalues $\gamma_1 \geq \dots \geq \gamma_d$ and $\boldsymbol{\Sigma}^{(j)}$ is a diagonal matrix with the diagonal terms $\sqrt{p_i} = \sqrt{\left[\sqrt{\frac{w_i}{\gamma_i}} - \frac{1}{\gamma_i} \right]^+}$.

V. SUM-RATE MAXIMIZATION

The SRM problem for a number of users $K > 1$ is non-convex even when removing the constraints on the number of streams per user. The general problem in fact remains non-convex and is NP-hard [24]. Therefore, since finding the global optimal has prohibitive complexity, one needs to resort to suboptimal solutions with reasonable complexity. In this section, we review several suboptimal solutions to the SRM problem that have been proposed in the literature. Since some of these techniques were originally proposed for a scenario that does not subsume the considered MIMO-IFC-GC, we also propose the necessary modifications required for application to the MIMO-IFC-GC. Note that these techniques perform an optimization over the transmit covariance matrices by relaxing the rank constraint due to the number of users per streams (see discussion below). Therefore, we also review and modify when necessary a different class of algorithms that solve problems related to SRM but are able to enforce constraints on the number of transmitting streams per user. The WSMMSE problem does not seem to have been addressed previously for the MIMO-IFC-GC and will be studied in the next section.

A. Soft Interference Nulling

A solution to the SRM problem for the MIMO-IFC-GC was proposed in [12]. In this technique the optimization is over all transmit covariance matrices $\boldsymbol{\Sigma}_k = \mathbf{B}_k \mathbf{B}_k^H \in \mathbb{C}^{m_{t,k} \times m_{t,k}}$. The constraints on the number of streams would impose a rank constraint on $\boldsymbol{\Sigma}_k$ as $\text{rank}(\boldsymbol{\Sigma}_k) = d_k$. Here, and in all the following reviewed techniques below, unless stated otherwise, such rank constraints are relaxed by assuming that the number of transmitting data streams is equal to the transmitting antennas to that user, i.e. $d_k = m_{t,k}$. From (15) and (18), we can rewrite the (negative) sum-rate as

$$\begin{aligned} \sum_{k=1}^K \log |\mathbf{E}_k| &= - \sum_{k=1}^K \log |\boldsymbol{\Omega}_k + \mathbf{H}_{k,k} \boldsymbol{\Sigma}_k \mathbf{H}_{k,k}^H| \\ &\quad + \log |\boldsymbol{\Omega}_k|, \end{aligned} \quad (29)$$

where $\boldsymbol{\Omega}_k$ is defined in (12). Notice that it is often convenient to work with the covariance matrices instead of the beamforming matrices \mathbf{B}_k , since this change of variables may render the optimization problem convex as, for instance, when minimizing the first term only in (29). It can then be seen that the SRM problem is, however, non-convex due to the presence of the $-\log |\boldsymbol{\Omega}_k|$ term, which is indeed a concave function of the matrices $\boldsymbol{\Sigma}_k$.

An approximate solution is then found in [12] via an iterative scheme, whereby at each $(j+1)$ th iteration, given the previous solution $\boldsymbol{\Sigma}_k^{(j)}$ the non-convex term $-\log |\boldsymbol{\Omega}_k|$ is approximated using a first-order Taylor expansion as

$$\begin{aligned} -\log |\boldsymbol{\Omega}_k| &\simeq -\log |\boldsymbol{\Omega}_k^{(j)}| \\ &\quad - \sum_{l \neq k} \text{tr} \left\{ \left(\boldsymbol{\Omega}_k^{(j)} \right)^{-1} \mathbf{H}_{k,l} \left(\boldsymbol{\Sigma}_l - \boldsymbol{\Sigma}_l^{(j)} \right) \mathbf{H}_{k,l}^H \right\}, \end{aligned} \quad (30)$$

where $\Omega_k^{(j)} = \mathbf{I} + \sum_{l \neq k} \mathbf{H}_{k,l} \Sigma_k^{(j)} \mathbf{H}_{k,l}^H$. Since the resulting problem

$$\begin{aligned} \underset{\Sigma_k, k=1, \dots, K}{\text{minimize}} \quad & - \sum_{k=1}^K \log |\Omega_k + \mathbf{H}_{k,k} \Sigma_k \mathbf{H}_{k,k}| \\ & + \sum_{l \neq k} \text{tr} \left\{ \left(\Omega_k^{(j)} \right)^{-1} \mathbf{H}_{k,l} \Sigma_l \mathbf{H}_{k,l}^H \right\} \\ \text{subject to} \quad & \text{tr} \{ \Phi_{k,m} \Sigma_k \} \leq P_m, \quad m = 1, \dots, M, \end{aligned} \quad (31)$$

is convex, a solution can be found efficiently. Following the original reference [12], we refer to this scheme as “soft interference nulling”. We refer to [12] for further details about the algorithm.

B. SDP Relaxation

A related approach is taken in [13] for the SRM problem² for a MIMO-IFC with regular per-transmitter, rather than generalized, power constraints. Similarly to the previous technique, the optimization is over the transmit covariance matrices and under the relaxed rank constraints. In particular, the authors first approximate the problem by using the approach in [18]. Then, an iterative solution is proposed by linearizing a non-convex term similar to soft interference nulling as reviewed above. It turns out that such linearized problem can be solved using semi-definite programming (SDP). Specifically, denoting with $\Omega_k^{(j)}$ the matrix (12) corresponding to the solution $\mathbf{B}_k^{(j)}$ at the previous iteration j , i.e., $\Omega_k^{(j)} = \mathbf{I} + \sum_{l \neq k} \mathbf{H}_{k,l} \mathbf{B}_l^{(j)} \mathbf{B}_l^{(j)H} \mathbf{H}_{k,l}^H$, the SDP problem to be solved to find the solutions $\mathbf{B}_k^{(j+1)}$ for the $(j+1)$ th iteration is given by

$$\begin{aligned} \underset{\mathbf{Y}_k, \Sigma_k, \forall k}{\text{minimize}} \quad & \sum_{k=1}^K \text{tr} \{ \mathbf{Y}_k \} + \sum_{k=1}^K \text{tr} \left\{ \mathbf{C}_k^{(j)} \Sigma_k \right\} \\ \text{subject to} \quad & \sum_{k=1}^K \text{tr} \{ \Phi_{k,m} \Sigma_k \} \leq P_m, \quad m = 1, \dots, M \\ & \begin{bmatrix} \mathbf{H}_{k,k} \Sigma_k \mathbf{H}_{k,k}^H + \Omega_k^{(j)} \left(\mathbf{W}_k^{(j)} \Omega_k^{(j)} \right)^{\frac{1}{2}} \\ \left(\mathbf{W}_k^{(j)} \Omega_k^{(j)} \right)^{\frac{1}{2}} \mathbf{Y}_k \end{bmatrix} \succeq 0, \\ & \text{and } \Sigma_k \succeq 0, \quad k = 1, \dots, K \end{aligned}$$

where

$$\mathbf{W}_k^{(j)} = \mathbf{I} + \mathbf{H}_{k,k} \Sigma_k^{(j)} \mathbf{H}_{k,k}^H \Omega_k^{(j)-1}, \quad (32)$$

$$\begin{aligned} \mathbf{C}_k^{(j)} = \sum_{i \neq k} \mathbf{H}_{i,k}^H \left(\mathbf{I} + \sum_l \mathbf{H}_{i,l} \Sigma_l^{(j)} \mathbf{H}_{i,l}^H \right)^{-1} \mathbf{W}_i^{(j)} \times \\ \mathbf{H}_i \Sigma_i^{(j)} \mathbf{H}_i^H \left(\mathbf{I} + \sum_l \mathbf{H}_{i,l} \Sigma_l^{(j)} \mathbf{H}_{i,l}^H \right)^{-1} \mathbf{H}_{i,k}, \quad (33) \end{aligned}$$

and \mathbf{Y}_k is an auxiliary optimization variable, defined using the Schur complement as $\mathbf{Y}_k = \mathbf{W}_k \Omega_k^{(j)} \left(\mathbf{H}_{k,k} \Sigma_k \mathbf{H}_{k,k}^H + \Omega_k^{(j)} \right)^{-1}$ to convert the original optimization problem to an SDP problem [13]. The derivation requires minor modifications with respect to [13] and is

²More generally, the reference studies the weighted SRM problem.

therefore not detailed. The scheme is referred to as “SDP relaxation” in the following. We refer to [13] for further details about the algorithm.

C. Polite Waterfilling

Reference [16] studied the (weighted) SRM problem for a general model that includes the MIMO-IFC-GC. We review the approach here for completeness. Two algorithms are proposed, whose basic idea is to search iteratively for a solution of the KKT conditions [11] for the (weighted) SRM problem. Notice that, since the problem is non-convex, being a solution of the KKT conditions is only necessary (as proved in [16]) but not sufficient to guarantee global optimality. It is shown in [16] that any solution Σ_k , $k = 1, \dots, K$, of the KKT conditions must have a specific structure that is referred to as “polite waterfilling”, which is reviewed below for the SRM problem.

Lemma 6 [16]: For a given set of Lagrange multipliers $\lambda = (\mu \lambda_1, \dots, \mu \lambda_M)$, where $\mu > 0$ and $\lambda_i \geq 0$ for $i = 1, \dots, M$, associated with the M power constraints in (13), define the covariance matrices

$$\hat{\Omega}_k = \sum_{m=1}^M \lambda_m \Phi_{k,m} + \sum_{j \neq k} \mathbf{H}_{j,k}^H \hat{\Sigma}_j \mathbf{H}_{j,k}, \quad (34)$$

with

$$\hat{\Sigma}_k = \frac{1}{\mu} \left(\Omega_k^{-1} - (\Omega_k + \mathbf{H}_{k,k} \Sigma_k \mathbf{H}_{k,k}^H)^{-1} \right). \quad (35)$$

An optimal solution Σ_k , $k = 1, \dots, K$, of the SRM problem must have the “polite waterfilling” form

$$\Sigma_k = \hat{\Omega}_k^{-\frac{1}{2}} \mathbf{V}_k \mathbf{P}_k \mathbf{V}_k^H \hat{\Omega}_k^{-\frac{1}{2}}, \quad (36)$$

where the columns of \mathbf{V}_k are the right singular vectors of the “pre- and post- whitened channel matrix” $\Omega_k^{-\frac{1}{2}} \mathbf{H}_{k,k} \hat{\Omega}_k^{-\frac{1}{2}}$ with (12) for $k = 1, \dots, K$, and \mathbf{P}_k is a diagonal matrix with diagonal elements $p_{k,i}$. The powers $p_{k,i}$ must satisfy

$$p_{k,i} = \left[\frac{1}{\mu} - \frac{1}{\gamma_{k,i}} \right]^+, \quad (37)$$

where $\sqrt{\gamma_{k,i}}$ is the i th singular value of the whitened matrix $\Omega_k^{-\frac{1}{2}} \mathbf{H}_{k,k} \hat{\Omega}_k^{-\frac{1}{2}}$. Parameter $\mu \geq 0$ is selected so as to satisfy the constraint

$$\sum_{m=1}^M \lambda_m \sum_{k=1}^K \text{tr} \{ \Phi_{k,m} \Sigma_k \} \leq \sum_{m=1}^M \lambda_m P_m, \quad (38)$$

which implied by the constraints of the original problem (13). Moreover, parameters $\lambda_i \geq 0$ are to be chosen so as to satisfy each individual constraint in (13).

In order to obtain a solution Σ_k , $k = 1, \dots, K$, according to polite waterfilling form as described in Lemma 6, [16] proposes to use the interpretation of $\hat{\Omega}_k$ in (34) as the interference plus noise covariance matrix and $\hat{\Sigma}_k$ in (35) as the transmit covariance matrix both at the “dual” system³.

³In the “dual” system the role of transmitters and receivers is switched, i.e., the k th transmitter in the original system becomes the k th receiver in the “dual” system. The channel matrix between the k th transmitter and the l th receiver in the dual system is given by $\mathbf{H}_{l,k}^H$.

Based on this observation, the algorithm proposed in [16] works as follows. At each j th iteration, first one calculates the covariance matrices $\Sigma_k^{(j)}$ in the original system using the polite waterfilling solution of Lemma 6; then one calculates the matrices $\hat{\Sigma}_k^{(j)}$ using again polite waterfilling in the dual system as explained above. Finally, at the end of each j th iteration, one updates the Lagrange multipliers as

$$\lambda_m^{(j+1)} = \lambda_m^{(j)} \frac{\sum_{k=1}^K \text{tr} \left\{ \Phi_{k,m} \Sigma_k^{(j)} \right\}}{P_m}, \quad (39)$$

thus forcing the solution to satisfy the constraints of the SRM problem (13). For details on the algorithm, we refer to [16].

Remark 2: Other notable algorithms designed to solve the SRM problem for the special case of a MIMO-BC with generalized constraints are [25], [26]. As explained in [16], these schemes are not easily generalized to the scenario at hand where the cost function is not convex. As such, they will not be further studied here.

D. Leakage Minimization

While the techniques discussed above do not enforce constraints on the number of stream per users, here we extend a technique previously proposed in [27] that aims at aligning interference through minimizing the interference leakage and is able to enforce the desired rank constraints. It is known that this approach is solves the SRM problem for high signal-to-noise-ratio (SNR). In this algorithm, it is assumed that the power budget is divided equally between the data streams and the precoding matrix of user k from BS m is given as $\mathbf{B}_{k,m} = \sqrt{\frac{P_m}{K_m d_k}} \bar{\mathbf{B}}_{k,m}$ where $\bar{\mathbf{B}}_{k,m}$ is a $n_t \times d_k$ matrix of orthonormal columns (i.e. $\bar{\mathbf{B}}_{k,m}^\text{H} \bar{\mathbf{B}}_{k,m} = \mathbf{I}$). The equalization matrices are also assumed to have orthonormal columns (i.e. $\mathbf{A}_k^\text{H} \mathbf{A}_k = \mathbf{I}$). Hence, there is no inter-stream interference for each user. Total interference leakage at user k is given by

$$I = \sum_k \text{tr} \left\{ \mathbf{A}_k^\text{H} \mathbf{Q}_k \mathbf{A}_k \right\}. \quad (40)$$

where $\mathbf{Q}_k = \sum_{j \neq k} \sum_{m \in \mathcal{M}_j} \frac{P_m}{K_m d_j} \tilde{\mathbf{H}}_{k,m} \bar{\mathbf{B}}_{k,m} \bar{\mathbf{B}}_{k,m}^\text{H} \tilde{\mathbf{H}}_{j,m}^\text{H}$. To minimize the interference leakage, the equalization matrix \mathbf{A}_k for user k can be obtained as $\mathbf{A}_k = v_{d_k}(\mathbf{Q}_k)$ where $v_{d_k}(\mathbf{A})$ represents a matrix with columns given by the eigenvectors corresponding to the d_k smallest eigenvalues of \mathbf{A} . Now, for fixed matrices \mathbf{A}_k , the cost function (40) can be rewritten as

$$I = \sum_k \sum_{m \in \mathcal{M}_k} \text{tr} \left\{ \bar{\mathbf{B}}_{k,m}^\text{H} \hat{\mathbf{Q}}_{k,m} \bar{\mathbf{B}}_{k,m} \right\} \quad (41)$$

where $\hat{\mathbf{Q}}_{k,m} = \sum_{j \neq k, j \in \mathcal{K}_m} \frac{P_m}{K_m d_k} \tilde{\mathbf{H}}_{j,m}^\text{H} \mathbf{A}_j \mathbf{A}_j^\text{H} \tilde{\mathbf{H}}_{j,m}$.⁴ Minimizing over the matrices \mathbf{B}_k leads to choosing $\bar{\mathbf{B}}_{k,m} = v_{d_k}(\hat{\mathbf{Q}}_{k,m})$. The algorithm iterates until convergence. We refer to this scheme as “min leakage” in the following.

⁴In the original work [27] which is proposed for the interference channels, the algorithm iteratively exchanges the role of transmitters and receivers to update the precoding and equalization matrices similarly.

E. Max-SINR

Another algorithm called “max-SINR” has been proposed in [27] which is based on the maximization of SINR, rather than directly the sum-rate. This algorithm is also able to enforce rank constraints. The max-SINR algorithm assumes equal power allocated to the data streams and attempts at maximizing the SINR for each stream by selecting the receive filters. Then, it exchanges the role of transmitter and receiver to obtain the transmit precoding matrices which maximizes the max-SINR. This iterates until convergence. A modification of this algorithm is given in [28] by maximizing the ratio of the average signal power to the interference plus noise power (SINR-like) term. However, these techniques are only given for standard MIMO interference channels and not for MIMO-IFC-GC.

VI. MSE MINIMIZATION

In this section, we propose two suboptimal techniques to solve the WSMMSE problem. We recall that with the WSMMSE problem enforcing the constraint on d_k is necessary in order to avoid trivial solutions. Performance comparison among all the considered schemes will be provided in Sec. VII for a multi-cell system with network MIMO.

A. MMSE Interference Alignment

A technique referred to as MMSE interference alignment (MMSE-IA) was presented in [19] for an interference channel with per-transmitter power constraints and where each receiver is endowed with a single antenna. Here we extend the approach to the MIMO-IFC-GC.

The idea is to approximate the solution of the WSMMSE problem by optimizing the precoding matrices \mathbf{B}_k followed by the equalization matrices \mathbf{A}_k and iterating the procedure. Specifically, initialize \mathbf{B}_k arbitrarily. Then, at each iteration j : (i) For each user k , evaluate the equalization matrices using the MMSE solution (18), obtaining $\mathbf{A}_k^{(j)} = (\mathbf{H}_{k,k} \mathbf{B}_k^{(j-1)} \mathbf{B}_k^{(j-1)\text{H}} \mathbf{H}_{k,k}^\text{H} + \boldsymbol{\Omega}_k^{(j-1)})^{-1} \mathbf{H}_{k,k} \mathbf{B}_k^{(j-1)}$, where from (12) we have $\boldsymbol{\Omega}_k^{(j-1)} = \mathbf{I} - \sum_{l \neq k} \mathbf{H}_{k,l} \mathbf{B}_l^{(j-1)} \mathbf{B}_l^{(j-1)\text{H}} \mathbf{H}_{k,l}^\text{H}$; (ii) Given the matrices $\mathbf{A}_k^{(j)}$, the WSMMSE problem becomes

$$\begin{aligned} & \underset{\mathbf{B}_k, k=1, \dots, K}{\text{minimize}} \quad \sum_{k=1}^K \text{tr} \left\{ \mathbf{W}_k \mathbf{E}_k^{(j)} \right\} \\ & \text{subject to} \quad \sum_{k=1}^K \text{tr} \left\{ \Phi_{k,m} \mathbf{B}_k \mathbf{B}_k^\text{H} \right\} \leq P_m, \forall m \in \mathcal{M} \end{aligned} \quad (42)$$

where $\mathbf{E}_k^{(j)}$ is (11) with $\mathbf{A}_k^{(j)}$ in place of \mathbf{A}_k . Fixing the equalization matrices $\mathbf{A}_k^{(j)}, \forall k$, this problem is convex in \mathbf{B}_k and can be solved by enforcing the KKT conditions. Therefore, matrices $\mathbf{B}_k^{(j)}$ for the j th iteration can be obtained as follows.

Lemma 7: For given equalization matrices $\mathbf{A}_k^{(j)}$, a solution $\mathbf{B}_k^{(j)}, k = 1, \dots, K$, of the WSMMSE problem is given by

$$\begin{aligned} \mathbf{B}_k^{(j)} = & \left(\sum_{l=1}^K \mathbf{H}_{l,k}^\text{H} \mathbf{A}_l^{(j)} \mathbf{W}_l \mathbf{A}_l^{(j)\text{H}} \mathbf{H}_{l,k} + \sum_m \mu_m \Phi_{k,m} \right)^{-1} \times \\ & \mathbf{H}_{k,k}^\text{H} \mathbf{A}_k^{(j)} \mathbf{W}_k \end{aligned} \quad (43)$$

where μ_m are Lagrangian multipliers satisfying

$$\mu_m \geq 0 \quad (44)$$

$$\mu_m \left(\sum_{k=1}^K \text{tr} \left\{ \Phi_{k,m} \mathbf{B}_k^{(j)} \mathbf{B}_k^{(j)\text{H}} \right\} - P_m \right) = 0 \quad (45)$$

and the power constraints $\sum_{k=1}^K \text{tr} \left\{ \Phi_{k,m} \mathbf{B}_k^{(j)} \mathbf{B}_k^{(j)\text{H}} \right\} \leq P_m$ for all m .

Once obtained the matrices $\mathbf{B}_k^{(j)}$ using the results in Lemma 7, the iterative procedure continues with the $(j+1)$ th iteration. We refer to this scheme as extended MMSE-IA, or eMMSE-IA.

Remark 3: The algorithm proposed above reduces to the one introduced in [19] in the special case of per-transmitter power constraints and single-antenna receivers. It is noted that in such case, problem (42) can be solved in a distributed fashion, so that each transmitter k can calculate its matrix (more precisely vector, given the single antenna at the receivers) independently from the other transmitters. In the MIMO-IFC-GC, the power constraints couple the solutions of the different users and thus make a distributed approach infeasible.

B. Diagonalized MMSE

We now propose an iterative optimization strategy inspired by the single-user algorithm that we put forth in Sec. IV. At the $(j+1)$ th iteration, given the matrices obtained at the previous iteration, we proceed as follows. The weighted sum-MSE (14) with the definition of MSE-matrices (11) is a convex function in each \mathbf{A}_k and \mathbf{B}_k when $(\mathbf{B}_j, \mathbf{A}_j), \forall j \neq k$ are fixed. Nevertheless, it is not jointly convex in terms of both $(\mathbf{A}_k, \mathbf{B}_k)$. Inspired by Lemma 5 for the corresponding single-user problem, we propose a (suboptimal) solution based on the solution of the dual problem for calculation of $(\mathbf{A}_k, \mathbf{B}_k)$. To this end, we first obtain \mathbf{A}_k as (18). Then, we simplify the Lagrangian function with respect to \mathbf{B}_k by removing the terms independent of \mathbf{B}_k . Specifically, by defining $\boldsymbol{\Upsilon}_k = \sum_{l \neq k} \mathbf{H}_{l,k}^H \mathbf{A}_l \mathbf{W}_l \mathbf{A}_l^H \mathbf{H}_{l,k}$, we have that the Lagrangian function at hand is given by

$$\begin{aligned} \mathcal{L}(\mathbf{B}_k; \boldsymbol{\lambda}) &= \text{tr} \left\{ \mathbf{W}_k \left(\mathbf{I} + \mathbf{B}_k^H \mathbf{H}_{k,k}^H \boldsymbol{\Omega}_k^{-1} \mathbf{H}_{k,k} \mathbf{B}_k \right)^{-1} \right\} \\ &\quad + \text{tr} \left\{ \boldsymbol{\Upsilon}_k \mathbf{B}_k \mathbf{B}_k^H \right\} \\ &\quad + \text{tr} \left\{ \left(\sum_m \lambda_m \Phi_{k,m} \right) \mathbf{B}_k \mathbf{B}_k^H \right\} \end{aligned} \quad (46)$$

This Lagrangian function for user k is the same as the Lagrangian function (25) of single-user WSMMSE problem when $\Phi(\boldsymbol{\lambda})$ is replaced with $\mathbf{F}_k(\boldsymbol{\lambda}) = \boldsymbol{\Upsilon}_k + \sum_m \lambda_m \Phi_{k,m}$. Matrix $\mathbf{F}_k(\boldsymbol{\lambda})$ is non-singular and therefore, using the same argument as in the proof of Lemma 5, for a given Lagrange multipliers $\boldsymbol{\lambda}$ and given other users' transmission strategies $(\mathbf{A}_l, \mathbf{B}_l), \forall l \neq k$, the optimal transmit precoding matrix can be obtained as

$$\mathbf{B}_k = \mathbf{F}_k(\boldsymbol{\lambda})^{-\frac{1}{2}} \mathbf{U}_k \boldsymbol{\Sigma}_k, \quad (47)$$

where $\mathbf{U}_k \in \mathbb{C}^{m_{t,k} \times d_k}$ is the eigenvectors of $\mathbf{F}_k(\boldsymbol{\lambda})^{-\frac{1}{2}} \mathbf{H}_{k,k}^H \boldsymbol{\Omega}_k^{-1} \mathbf{H}_{k,k} \mathbf{F}_k(\boldsymbol{\lambda})^{-\frac{1}{2}}$ corresponding to its

largest eigenvalues $\gamma_{k,1} \geq \dots \geq \gamma_{k,d_k}$ and $\boldsymbol{\Sigma}_k$ is diagonal matrices with the elements $\sqrt{p_{k,i}}$ given by

$$p_{k,i} = \left[\sqrt{\frac{w_{k,i}}{\gamma_{k,i}}} - \frac{1}{\gamma_{k,i}} \right]^+, \quad (48)$$

with $\boldsymbol{\lambda} \succeq 0$ being the Lagrangian multipliers satisfy the power constraints. Since this scheme diagonalizes the MSE matrices defined in (9), it is referred to as diagonalized MMSE (DMMSE).

To summarize, the proposed algorithm at each iteration j (i) evaluates the transmit precoding matrices $\mathbf{B}_k^{(j)}$ given other users' transmission strategies $(\mathbf{A}_l^{(j-1)}, \mathbf{B}_l^{(j-1)})$ using (47)-(48) (ii) updates the equalization matrices using the MMSE solution (18); (iii) updates the $\boldsymbol{\lambda}$ via a subgradient update

$$\lambda_m^{(j+1)} = \lambda_m^{(j)} + \delta \left(P_m - \sum_{k=1}^K \text{tr} \left\{ \Phi_{k,m} \mathbf{B}_k \mathbf{B}_k^H \right\} \right) \quad (49)$$

to satisfy the power constraints.

Remark 4: In this paper, we assume perfect knowledge of channel state information (CSI). Therefore, each transmitter and receiver has sufficient information to calculate the resulting precoders and equalizers by running the proposed algorithms. Under this assumption, which is common to other reviewed works such as [12][13], no exchange of precoder and equalizer vectors is required between the transmitters and receivers. Nevertheless, in practice, the CSI may only be available locally, in the sense that transmitter k knows channel matrices $\mathbf{H}_{l,k}$, for all $l = 1, \dots, K$, whereas receiver k is aware of channel matrices $\mathbf{H}_{k,l}$, for all $l = 1, \dots, K$. The proposed DMMSE and the reviewed PWF [14][16] algorithms require, beside the local CSI, that the transmitter k has available also the interference plus noise covariance matrix, $\boldsymbol{\Omega}_k$, and the current equalization matrices \mathbf{A}_l for all $l = 1, \dots, K$ in order to update the precoder for user k . Hence, to enable DMMSE and PWF with local CSI, exchange of the equalizer matrices is needed between the nodes. Similarly, the proposed eMMSEIA, and min leakage and Max-SINR algorithms [27], require the transmitters to know the equalizing matrices \mathbf{A}_l for $l = 1, \dots, K$ at each iteration, in addition to the local CSI. Moreover, each receiver must know the current precoders \mathbf{B}_l for all $l = 1, \dots, K$. Therefore, the overhead for the proposed eMMSEIA and the min leakage and Max-SINR algorithms involves the exchange of equalizer and precoder matrices between the transmitters and receivers. However, these latter algorithms can also be adapted using the bi-directional optimization process proposed in [29]. This process involves bi-directional training followed by data transmission. In the forward direction, the training sequences are sent using the current precoders. Then, at the user receivers the equalizers are updated to minimize the least square error cost function. In the backward training phase, the current equalizers are used to send the training sequences and the precoders are updated accordingly. Finally, the SIN [12] and SDP relaxation [13] techniques are applied in a centralized fashion (rather than by updating the transmitter and receiver for each user at each iteration), and they require centralized full knowledge of all channel matrices.

Remark 5: Reference [13] addresses the SRM problem for a MIMO-IFC with regular per-transmitter, rather than generalized, power constraints. The problem is addressed by solving an SDP problem at each iteration. Moreover, the optimization is over the transmit covariance matrices and under the relaxed rank constraint. This enforces a constraint on the number of transmitted streams per user. References [14]-[16] study the (weighted) SRM problem by decomposing the multiuser problem into single-user problems for each user. Each single-user problem is a standard single-user SRM problem with an additional interference power constraint. The approach used in [14]-[16] assumes that the number of transmitted streams is equal to n_r . In this paper, we address WSMMSE problem and allow for an arbitrary number of streams ($d_k \leq n_r$).

Remark 6: Our algorithms consists of an inner loop, which solves the WSMMSE problem, and an outer loop, which is the subgradient algorithm to update λ . The subgradient algorithm in the outer loop is convergent (with a proper selection of the step sizes [30]) due to the fact that the dual function $\inf_{\mathbf{B}} \mathcal{L}(\mathbf{B}; \lambda)$ is a concave function with respect to λ [11]. The inner loops of the proposed algorithms in this paper (i.e. eMMSEIA and DMMSE) are convergent since the objective function decreases at each iteration. A discussion of the convergence for a special case of the eMMSEIA algorithm can be found in [19]. However, the original problem is non-convex and our solutions are only local minima. Nevertheless, the DMMSE algorithm is shown to converge to a local minimum with better performance compared to the previously known schemes in Sec. VII.

VII. NUMERICAL RESULTS

We consider a hexagonal cellular system where each BS is equipped with n_t transmit antennas and each user has n_r receive antennas. The users are located uniformly at random. Two tiers of surrounding cells are considered as interference for each cluster. We consider the worst-case scenario for the inter-cluster interference, which will be the condition that interfering BSs transmit at the full allowed power [7], [8], [31], [32]. We define the cooperation factor κ as a number of BSs cooperating on transmission to each user. The κ BSs are assigned to each user so that the corresponding channel norms (or, alternatively, the corresponding received SNRs) are the largest.

The propagation channel between each BS's transmit antennas and mobile user's receive antenna is characterized by path loss, shadowing and Rayleigh fading. The path loss component is proportional to $d_{km}^{-\beta}$, where d_{km} denotes distance from base station m to mobile user k and $\beta = 3.8$ is the path loss exponent. The channel from the transmit antenna t of the base station b at the receive antenna r of the k th user is given by [7]

$$\mathbf{H}_{k,b}^{(r,t)} = \alpha_{k,b}^{(r,t)} \sqrt{\gamma_0 \rho_{k,b} A \left(\Theta_{k,b}^{(t)} \right) \left(\frac{d_{k,b}}{d_0} \right)^{-\beta}} \quad (50)$$

where $\alpha_{k,b}^{(r,t)} \sim \mathcal{CN}(0, 1)$ represents Rayleigh fading, $\rho_{k,b}^{(\text{dBm})}$ is the lognormal shadow fading between b th BS and k th user with standard deviation of 8 dB, and $d_0 = 1$ km is the cell

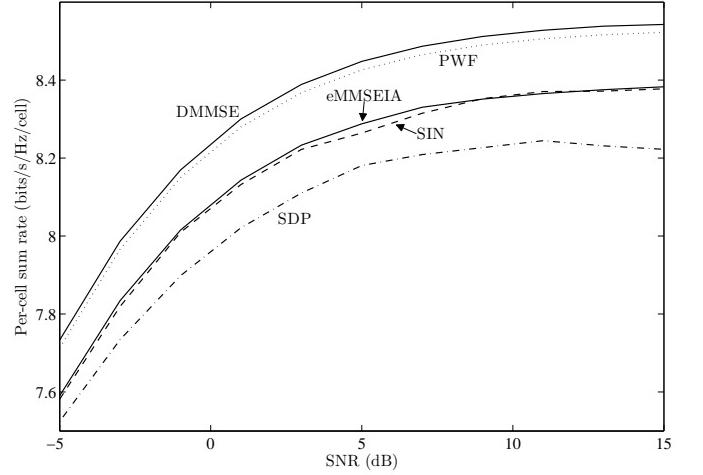


Fig. 2. Per-cell sum-rate for a MIMO-IFC-GC with $M = 3$ and $\kappa = 2$.

radius. γ_0 is the interference-free SNR at the cell boundary. We consider one user randomly located per cell for the numerical results.

When sectorization is employed, the transmit antennas are equally divided among the sectors of a cell. Each transmit antenna has a parabolic beam pattern as a function of the direction of the user from the broadside direction of the antenna (For more details refer to [7], [33]). The antenna gain is a function of the direction of the user k from the broadside direction of the t th transmit antenna of the b th base station denoted by $\Theta_{k,b}^{(t)} \in [-\pi, \pi]$; $\Theta_{3\text{dB}}$ is the half-power angle and A_s is the sidelobe gain. The antenna gain is given as [33]

$$A \left(\Theta_{k,b}^{(t)} \right)_{\text{dB}} = -\min \left(12 \left(\frac{\Theta_{k,b}^{(t)}}{\Theta_{3\text{dB}}} \right)^2, A_s \right) \quad (51)$$

For the 3,6-sector cells $A_s = 20, 23$ dB and $\Theta_{3\text{dB}} = \frac{70\pi}{180}, \frac{35\pi}{180}$, respectively [7], [33], [34]. When there is no sectorization we set $A = 1$.

We first compare different algorithms (for the solution of the SRM problem) without enforcing rank constraints on SIN, PWF, SDP relaxation and setting $d_k = \min(m_{t,k}, m_{r,k}) = n_r$ for the eMMSEIA and DMMSE algorithms. To solve the SRM problem, the weight matrices in the eMMSEIA and DMMSE algorithms are updated at each iteration as $\mathbf{W}_k = \mathbf{E}_k^{-1}$ using the current MSE-matrix \mathbf{E}_k . Fig. 2 compares the per-cell sum-rate of the algorithms discussed in this paper for a cluster with $M = 3$ cells and a cooperation factor $\kappa = 2$. The results show that our proposed DMMSE algorithm outperforms other techniques, while the polite water-filling algorithm (PWF) [14], [16] has a similar performance. Our proposed eMMSEIA scheme converges to a poorer local optimum value compared to these two schemes. The soft interference nulling (SIN) [12] and SDP relaxation [13] algorithms, which use the approximation of the non-convex terms in the objective function, perform worse in this example.

In Fig. 3, we evaluate the effect of partial cooperation for the DMMSE, eMMSEIA, and PWF algorithms in a cluster of

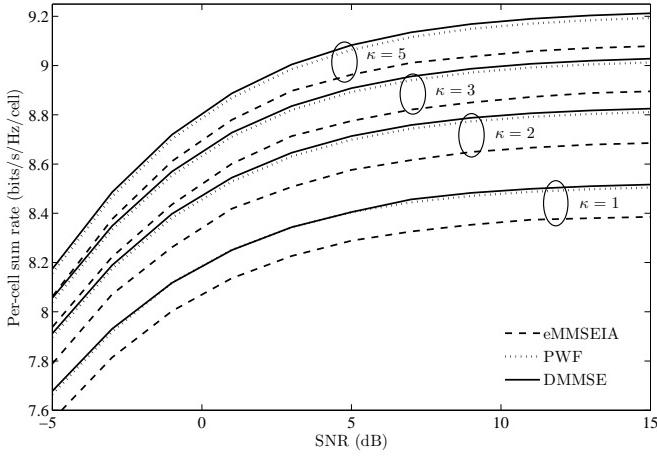


Fig. 3. Per-cell sum-rate for a MIMO-IFC-GC with $M = 5$ and $\kappa = 1, 2, 3, 5$, $n_t = 4$, $n_r = d_k = 2$, and 2 users per cell.

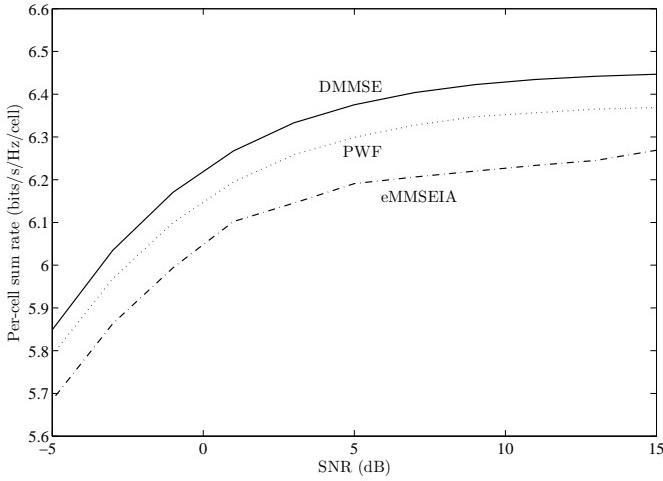


Fig. 4. Per-cell sum rate of the schemes that can support $d_k < \min(m_{t,k}, m_{r,k})$ for $d_k = 1$, $n_t = 4$, $n_r = 2$, $M = 3$ and $\kappa = 2$.

size $M = 5$ where each BS is equipped with $n_t = 4$ transmit antennas, each user employs $n_r = 2$ receive antennas, and 2 users are dropped randomly in each cell. Recall that the cooperation factor κ represents the number of BSs cooperating in transmission to each user. It can be seen that as κ increases the performance improves with diminishing returns as κ grows large. Moreover, the relative performance of the algorithms confirms the considerations above.

In Fig. 4, we compare again the performance of the schemes considered in Fig. 3 but with a stricter requirement on the number of streams, namely $d_k = 1$. It can be seen that the proposed DMMSE tends to perform better than PWF, which was not designed to handle rank constraints. We have adopted the PWF algorithm to support $d_k < \min(m_{t,k}, m_{r,k})$ by using a thin SVD of $\hat{\Omega}_k^{-\frac{1}{2}} \mathbf{H}_{k,k}^H \Omega_k^{-\frac{1}{2}}$ when computing (36).

In Fig. 5, we vary the size of the cluster M , showing also the advantages of coordinating transmission over larger clusters, even when the number of cooperating BSs κ is fixed.

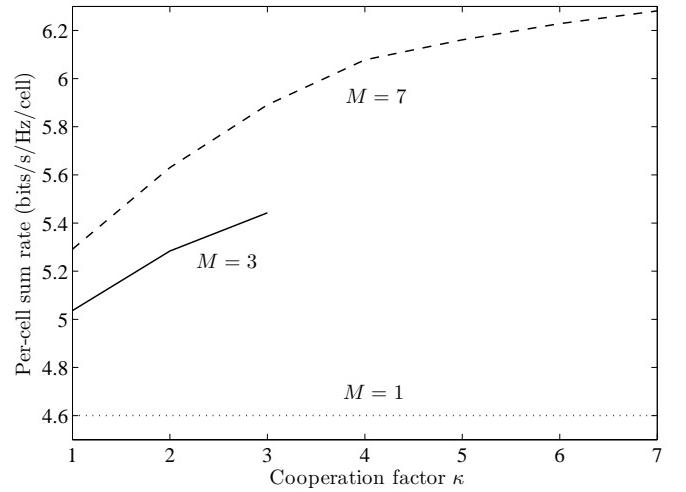


Fig. 5. Per-cell sum-rate of the proposed DMMSE scheme for cluster sizes $M = 1, 3, 7$ versus the cooperation factor, κ , with $n_t = n_r = 2$, SNR=20 dB, and single-user per cell.

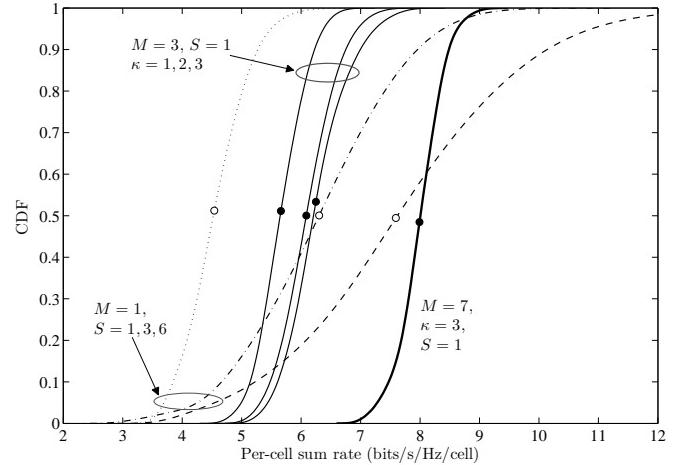


Fig. 6. CDF of the per-cell sum rates achieved by DMMSE for $S = 1, 3, 6$ sectors per cell, $M = 1, 3, 7$ coordinated clusters, and $\kappa = 1, 2, 3$ cooperation factors with $\gamma_0 = 20$ dB, $n_t = 6$, and $n_r = 2$. The circles represent the mean values of the per-cell sum-rates.

Recall that M represents the set of BSs whose transmission is coordinated, but only κ BSs cooperate for transmission to a given user. As an example, for a cluster size of $M = 7$ a cooperation factor of $\kappa = 4$ performs almost as well as the full cooperation scenario with $\kappa = 7$. Moreover, the performance gains with respect to the non-cooperative case $\kappa = 1$ are evident. We also show the performance with a cluster containing a single cell, i.e., $M = 1$. This highlights the performance gains attained even in the absence of message sharing among the BSs (i.e., $\kappa = 1$) due to the coordination of the BSs within the cluster.

Finally, the effect of sectorization is studied in Fig. 6 where $n_t = 6$ transmit antennas at each BS are divided equally into $S = 1, 3, 6$ sectors. Each cell contains 6 users, each equipped with $n_r = 2$ receive antennas. The users are randomly located

at the distance of $\frac{2}{3}d_0$ from its BS. For a given channel realization the DMMSE algorithm is used to obtain the per-cell sum rate. The cumulative distribution functions (CDFs) of per-cell sum rates are computed using large number of channel realizations. The gains of sectorization and cooperation are compared. For example, the system with coordination of 7 cells and $\kappa = 3$ cooperation factor and without sectorization performs better than the sectorized system with $S = 6$ and without any coordination between the BSs.

VIII. CONCLUSIONS

In this paper, we have studied a MIMO interference channel with partial cooperation at the BSs and per-BS power constraints. We have shown that the channel at hand is equivalent to a MIMO interference channel under generalized linear constraints (MIMO-IFC-GC). Focusing on linear transmission strategies, we have reviewed some of the available techniques for the maximization of the sum-rate and extended them to the MIMO-IFC-GC when necessary. Moreover, we have proposed two novel strategies for minimization of the weighted mean square error on the data estimates. Specifically, we have proposed an extension of the recently introduced MMSE interference alignment strategy and a novel strategy termed diagonalized MSE-matrix (DMMSE). Our proposed strategies support transmission of any arbitrary number of data streams per user. Extensive numerical results show that the DMMSE outperforms most previously proposed techniques and performs just as well as the best known strategy. Moreover, our results bring insight into the advantages of partial cooperation and sectorization and the impact of the size of the cooperating cluster of BSs and sectorization.

We conclude with a brief discussion on the complexity of the algorithms. Due to the difficulty of complete complexity analysis, especially in terms of speed of convergence, we present a discussion based on our simulation experiments. The PWF algorithm converges in almost the same number of iterations as the DMMSE algorithm. The complexity per iteration of PWF and DMMSE is also almost the same as $K(\mathcal{O}(\kappa n_t n_r^2) + \mathcal{O}(n_r^3))$ (required for the thin SVD operation). However, the PWF algorithm contains additional operations (matrix inversion and SVD) to obtain the precoding matrices from the calculated transmit covariance matrices.⁵ Also, the PWF algorithm includes a water-filling algorithm within its inner loop, which is not required in the DMMSE algorithm. The eMMSEIA algorithm has lower complexity per iteration (i.e. $K\mathcal{O}(n_r^3)$) than the PWF and DMMSE algorithms, since its complexity is due to a matrix inversion per iteration per user. However, eMMSEIA converges in a larger number of iterations than DMMSE and PWF. The complexity per iteration for the SDP relaxation is higher than for the SIN algorithm (this is because of the extra auxiliary positive semi-definite matrix variable, \mathbf{Y} , introduced in the SDP relaxation algorithm). The SIN algorithm also converges in a smaller number of iterations than the SDP relaxation algorithm.

⁵This can be performed together with finding the MMSE receive matrices.

APPENDIX PROOF OF LEMMA 5

The inequality (24) follows from weak Lagrangian duality. We now prove the second part of the statement. Recognizing now that $\text{tr}\{\mathbf{WE}\}$ with (19) is a Schur-concave function of the diagonal elements of (19)⁶, it can be argued that the minimum is attained when \mathbf{E} is diagonalized as we did for Lemma 4. Defining $\mathbf{R} = \mathbf{H}^H \mathbf{\Omega}^{-1} \mathbf{H}$, we can conclude that $\tilde{\mathbf{B}}^H \mathbf{R} \tilde{\mathbf{B}}$ must be also diagonal in this search domain. Now assume that an optimal solution of the single-user WSMSE problem is denoted as $\tilde{\mathbf{B}}$. Without loss of generality we can assume that this solution diagonalizes the MSE matrices. The necessity of the KKT conditions can be proved as in [16] and in special cases such as the MIMO interference channel with partial message sharing of Sec. II, it also follows from linear independence constraint qualification conditions [30].

Hence, there exists a Lagrange multiplier vector $\tilde{\boldsymbol{\lambda}}$ which together with $\tilde{\mathbf{B}}$ satisfies the KKT conditions of the WSMSE problem (20) [18][30]. As it is stated in the Lemma, we consider the case that $\tilde{\lambda}_m$ are also strictly positive (i.e. $\tilde{\lambda}_m > 0$ for all m). Simplifying the KKT condition (26), we have⁷

$$\nabla_{\mathbf{B}} \mathcal{L} = -\tilde{\mathbf{R}} \tilde{\mathbf{B}} \tilde{\mathbf{E}} \tilde{\mathbf{W}} \tilde{\mathbf{E}} + \sum_{m=1}^M \tilde{\lambda}_m \Phi_m \tilde{\mathbf{B}} = \mathbf{0} \quad (52)$$

Left-multiplying (52) by $\tilde{\mathbf{B}}^H$ gives us

$$\tilde{\mathbf{B}}^H \tilde{\mathbf{R}} \tilde{\mathbf{B}} \tilde{\mathbf{E}} \tilde{\mathbf{W}} \tilde{\mathbf{E}} = \tilde{\mathbf{B}}^H \left(\sum_m \tilde{\lambda}_m \Phi_m \right) \tilde{\mathbf{B}}. \quad (53)$$

Since $\tilde{\mathbf{B}}^H \tilde{\mathbf{R}} \tilde{\mathbf{B}}$ and correspondingly $\tilde{\mathbf{E}}$ are diagonal matrices, $\tilde{\mathbf{B}}^H \left(\sum_m \tilde{\lambda}_m \Phi_m \right) \tilde{\mathbf{B}}$ must also be diagonal. For simplicity, we introduce $\Phi(\tilde{\boldsymbol{\lambda}}) = \sum_{m=1}^M \tilde{\lambda}_m \Phi_m$. Since $\tilde{\lambda}_m > 0$ for every m , therefore $\Phi(\tilde{\boldsymbol{\lambda}})$ is a non-singular matrix. This can be easily verified due to the structure of Φ_m . Hence, we can write $\tilde{\mathbf{B}}^H \Phi(\tilde{\boldsymbol{\lambda}}) \tilde{\mathbf{B}} = \tilde{\Delta}$ where $\tilde{\Delta} \in \mathbb{C}^{d \times d}$ is a diagonal matrix. Therefore, we can write

$$\Phi(\tilde{\boldsymbol{\lambda}})^{1/2} \tilde{\mathbf{B}} = \tilde{\mathbf{U}} \tilde{\Sigma} \quad (54)$$

where $\tilde{\mathbf{U}} \in \mathbb{C}^{m_t \times d}$ consists of orthonormal columns (i.e. $\tilde{\mathbf{U}}^H \tilde{\mathbf{U}} = \mathbf{I}$) and $\tilde{\Sigma} \in \mathbb{C}^{d \times d}$ is a diagonal matrix with the diagonal terms of $\sqrt{\tilde{\lambda}_i}$. Hence, we can write

$$\tilde{\mathbf{B}} = \Phi(\tilde{\boldsymbol{\lambda}})^{-1/2} \tilde{\mathbf{U}} \tilde{\Sigma}. \quad (55)$$

Replacing the structure of $\tilde{\mathbf{B}}$ given in (55), we can write

$$\tilde{\mathbf{B}}^H \tilde{\mathbf{R}} \tilde{\mathbf{B}} = \tilde{\Sigma}^H \tilde{\mathbf{U}}^H \Phi(\tilde{\boldsymbol{\lambda}})^{-\frac{1}{2}} \mathbf{R} \Phi(\tilde{\boldsymbol{\lambda}})^{-\frac{1}{2}} \tilde{\mathbf{U}} \tilde{\Sigma} = \mathbf{D} \quad (56)$$

Thus, we can conclude from the equation above that $\tilde{\mathbf{U}}$ must contain the eigenvectors of $\Phi(\tilde{\boldsymbol{\lambda}})^{-\frac{1}{2}} \mathbf{R} \Phi(\tilde{\boldsymbol{\lambda}})^{-\frac{1}{2}}$.

⁶A Schur-concave function $f(\mathbf{x})$ of vector $\mathbf{x} = (x_1, \dots, x_d)$ is such that $f(\mathbf{x}) \leq f(\mathbf{x}')$ if \mathbf{x} majorizes \mathbf{x}' , that is, if $\sum_{i=1}^j x_{[i]} \geq \sum_{i=1}^j x'_{[i]}$ for all $j = 1, \dots, d$, where $x_{[i]}$ (and $x'_{[i]}$) represents the vector sorted in decreasing order, i.e., $x_{[1]} \geq \dots \geq x_{[d]}$ (and $x'_{[1]} \geq \dots \geq x'_{[d]}$).

⁷We use differentiation rule $\nabla_{\mathbf{X}} \text{tr}\{\mathbf{A} \mathbf{X}^H \mathbf{B}\} = \mathbf{B} \mathbf{A}$ and $\nabla_{\mathbf{X}} \text{tr}\{\mathbf{Y}^{-1}\} = -\mathbf{Y}^{-1} (\nabla_{\mathbf{X}} \mathbf{Y}) \mathbf{Y}^{-1}$. For the complex gradient operator each matrix and its conjugate transpose are treated as independent variables [35].

Now, plugging (55) into (26) and left-multiply it with $\Phi^{-\frac{1}{2}}$, we get

$$\tilde{\Gamma}\tilde{\Sigma}\left(\mathbf{I} + \tilde{\Gamma}\tilde{\Sigma}^2\right)^{-1}\mathbf{W}\left(\mathbf{I} + \tilde{\Gamma}\tilde{\Sigma}^2\right)^{-1} = \tilde{\Sigma} \quad (57)$$

where $\tilde{\Gamma}(\tilde{\lambda}) = \text{diag}[\gamma_1(\tilde{\lambda}) \cdots \gamma_d(\tilde{\lambda})]$ is a diagonal matrix with the diagonal terms of the d largest eigenvalues of $\Phi(\tilde{\lambda})^{-\frac{1}{2}}\mathbf{R}\Phi(\tilde{\lambda})^{-\frac{1}{2}}$. Since all the matrices are diagonal, (57) reduces to the scalar equations:

$$\frac{w_i\gamma_i(\tilde{\lambda})}{(1+\tilde{p}_i\gamma_i(\tilde{\lambda}))^2} = 1 \quad (58)$$

Solving these equations gives us the optimal \tilde{p}_i given by

$$\tilde{p}_i = \left[\sqrt{\frac{w_i}{\gamma_i(\tilde{\lambda})}} - \frac{1}{\gamma_i(\tilde{\lambda})} \right]^+, \quad (59)$$

Thus, for the given Lagrange multiplier $\tilde{\lambda}$ which together with $\tilde{\mathbf{B}}$, satisfying the KKT conditions of (20), $\tilde{\mathbf{B}}$ must satisfy (55) and (59). If all power constraints are satisfied with equality by this solution, then (55) and (59) also solves the single constraint problem

$$\begin{aligned} & \underset{\mathbf{B}}{\text{minimize}} \quad \text{tr} \left\{ \mathbf{W} (\mathbf{I} + \mathbf{B}^H \mathbf{H}^H \Omega^{-1} \mathbf{H} \mathbf{B})^{-1} \right\} \\ & \text{subject to} \quad \text{tr} \left\{ \Phi(\tilde{\lambda}) \mathbf{B} \mathbf{B}^H \right\} \leq \sum_{m=1}^M \tilde{\lambda}_m P_m, \end{aligned} \quad (60)$$

The solution of this problem is given in Lemma 3 as

$$\mathbf{B}(\tilde{\lambda}) = \Phi(\tilde{\lambda})^{-\frac{1}{2}} \mathbf{U} \Sigma \quad (61)$$

where \mathbf{U} consists of d eigenvectors of $\Phi(\tilde{\lambda})^{-\frac{1}{2}}\mathbf{R}\Phi(\tilde{\lambda})^{-\frac{1}{2}}$ corresponding to its largest eigenvalues and Σ is a diagonal matrix with the diagonal elements of $\sqrt{p_i}$, which is given by

$$p_{k,i} = \left[\sqrt{\frac{w_i}{\mu\gamma_i(\tilde{\lambda})}} - \frac{1}{\gamma_i(\tilde{\lambda})} \right]^+, \quad (62)$$

for a waterfilling value of $\mu \geq 0$ which satisfies the power constraint

$$\text{tr} \left\{ \Phi(\tilde{\lambda}) \mathbf{B}(\tilde{\lambda}) \mathbf{B}(\tilde{\lambda})^H \right\} \leq \sum_m \tilde{\lambda}_m P_m. \quad (63)$$

On the other hand, summing up the KKT conditions $\tilde{\lambda}_m (P_m - \text{tr} \left\{ \Phi_m \mathbf{B} \mathbf{B}^H \right\}) = 0$ for all m , we obtain that

$$\text{tr} \left\{ \left(\sum_m \tilde{\lambda}_m \Phi_m \right) \tilde{\mathbf{B}} \tilde{\mathbf{B}}^H \right\} = \sum_m \tilde{\lambda}_m P_m \quad (64)$$

If we set $\mu = 1$ and comparing (59) and (62), we can conclude that $\tilde{p}_i = p_i, \forall i$ which together with comparison of (61) and (55) we can conclude that $\mathbf{B}(\tilde{\lambda}) = \tilde{\mathbf{B}}$ and the $\mu = 1$ is the optimal Lagrange multiplier of the single-constraint WSMMSE problem (60). Following Lemma 4, this precoding matrix is also a result of minimization of the Lagrangian function (23) when $\mu = 1$ and $\Phi_1 = \Phi(\tilde{\lambda})$, which means

$$p^* = \inf_{\mathbf{B}} \mathcal{L}(\mathbf{B}; \tilde{\lambda}). \quad (65)$$

On the other hand, we have

$$\max_{\lambda \geq 0} \inf_{\mathbf{B}} \mathcal{L}(\mathbf{B}; \lambda) \geq \inf_{\mathbf{B}} \mathcal{L}(\mathbf{B}; \tilde{\lambda}) \quad (66)$$

which in concert with (24) and (65) results in

$$p^* = \inf_{\mathbf{B}} \mathcal{L}(\mathbf{B}; \tilde{\lambda}) = \max_{\lambda \geq 0} \inf_{\mathbf{B}} \mathcal{L}(\mathbf{B}; \lambda), \quad (67)$$

thus concluding the proof.

REFERENCES

- [1] H. Dai, A. Molisch, and H. Poor, "Downlink capacity of interference-limited MIMO systems with joint detection," *IEEE Trans. Wireless Commun.*, vol. 3, no. 2, pp. 442 – 53, Mar. 2004.
- [2] R. Blum, "MIMO capacity with interference," *IEEE J. Select. Areas Commun.*, vol. 21, no. 5, pp. 793 – 801, Jun. 2003.
- [3] D. Gesbert, S. Hanly, H. Huang, S. Shamai (Shitz), O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Select. Areas Commun.*, vol. 28, no. 9, pp. 1 – 29, Dec. 2010.
- [4] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 3936 – 3964, Sep. 2006.
- [5] V. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the K-user interference channel," *IEEE Trans. Inform. Theory*, vol. 54, no. 8, pp. 3425 – 3441, Aug. 2008.
- [6] M. Maddah-Ali, A. Motahari, and A. Khandani, "Communication over MIMO X channels: Interference alignment, decomposition, and performance analysis," *IEEE Trans. Inform. Theory*, vol. 54, no. 8, pp. 3457 – 3470, Aug. 2008.
- [7] H. Huang, M. Trivellato, A. Hottinen, M. Shafi, P. J. Smith, and R. Valenzuela, "Increasing downlink cellular throughput with limited network MIMO coordination," *IEEE Trans. Wireless Commun.*, vol. 8, no. 6, pp. 2983 – 2989, Jun. 2009.
- [8] J. Zhang, R. Chen, J. Andrews, A. Ghosh, and R. Heath, "Networked MIMO with clustered linear precoding," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 1910 – 1921, Apr. 2009.
- [9] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195 – 202, Jan. 2005.
- [10] R. R. Müller, D. Guo, and A. L. Moustakas, "Vector precoding for wireless MIMO systems and its replica analysis," *IEEE J. Select. Areas Commun.*, vol. 26, no. 3, pp. 530 – 540, Apr. 2008.
- [11] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge Univ. Press, 2004.
- [12] C. Ng and H. Huang, "Linear precoding in cooperative MIMO cellular networks with limited coordination clusters," *IEEE J. Select. Areas Commun.*, vol. 28, no. 9, pp. 1446–1454, Dec. 2010.
- [13] M. Razaviyayn, M. Sanjabi, and Z.-Q. Luo, "Linear transceiver design for interference alignment: Complexity and computation," submitted to the *IEEE Trans. on Inform. Theory*, Sep. 2010, preprint: arXiv:1009.3481v1.
- [14] A. Liu, Y. Liu, H. Xiang, and W. Luo, "Duality, polite water-filling, and optimization for MIMO B-MAC interference networks and iTree networks," submitted to the *IEEE Trans. Inform. Theory*, Apr. 2010, preprint: arXiv:1004.2484v2.
- [15] S. Kaviani, O. Simeone, W. A. Krzymień, and S. Shamai (Shitz), "Linear MMSE precoding and equalization for network MIMO with partial cooperation," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Dec. 2011.
- [16] A. Liu, Y. Liu, H. Xiang, and W. Luo, "Polite water-filling for weighted sum-rate maximization in B-MAC networks under multiple linear constraints," in *Proc. Allerton Conference on Commun., Control, and Comput.*, Sep. 2010.
- [17] D. P. Palomar and Y. Jiang, "MIMO transceiver design via majorization theory," *Found. Trends Commun. Inf. Theory* (NOW Publishers), vol. 3, no. 4 - 5, pp. 331-551, 2006.
- [18] S. Christensen, R. Agarwal, E. Carvalho, and J. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792–4799, Dec. 2008.
- [19] D. A. Schmidt, C. Shi, A. A. Berry, M. L. Honig, and W. Utschick, "Minimum mean squared error interference alignment," in *Proc. Asilomar Conf. on Signals, Systems and computers*, Nov. 2009.
- [20] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Processing*, vol. 59, no. 9, pp. 4331–4340, Sep. 2011.

- [21] H. Yu and V. K. N. Lau, "Rank-constrained schur-convex optimization with multiple trace/log-det constraints," *IEEE Trans. Signal Processing*, vol. 59, no. 1, pp. 304 – 314, Jan. 2011.
- [22] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: a unified framework for convex optimization," *IEEE Trans. Signal Processing*, vol. 51, no. 9, pp. 2381 – 2401, Sep. 2003.
- [23] H. Huh, G. Caire, S. H. Moon, Y.-T. Kim, and I. Lee, "Multi-cell MIMO downlink with cell cooperation and fair scheduling: a large-system limit analysis," submitted to the *IEEE Trans. Inform. Theory*, Jun. 2010, preprint: arXiv:1006.2162v1.
- [24] Z.-Q. Luo and S. Zhang, "Dynamic spectrum management: Complexity and duality," *IEEE J. Select. Areas Commun.*, vol. 2, no. 1, pp. 57 – 73, Feb. 2008.
- [25] H. Huh, H. Papadopoulos, and G. Caire, "MIMO broadcast channel optimization under general linear constraints," in *Proc. IEEE Int. Symp. on Infor. Theory (ISIT)*, Jul. 2009, pp. 2664 – 2668.
- [26] H. Huh, H. C. Papadopoulos, and G. Caire, "Multiuser MISO transmitter optimization for intercell interference mitigation," *IEEE Trans. Signal Processing*, vol. 58, no. 8, pp. 4272 –4285, Aug. 2010.
- [27] K. Gomadam, V. Cadambe, and S. A. Jafar, "Approaching the capacity of wireless networks through distributed interference alignment," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Nov. 2008.
- [28] S. W. Peters and R. W. Heath, "Cooperative algorithms for MIMO interference channels," *IEEE Trans. Veh. Technol.*, vol. 60, no. 1, pp. 206 –218, Jan. 2011.
- [29] C. Shi, R. Berry, and M. Honig, "Adaptive beamforming in interference networks via bi-directional training," in *Proc. 44th Annual Conference on Information Sciences and Systems (CISS)*, Mar. 2010.
- [30] D. P. Bertsekas, A. Nedić, and A. E. Ozdaglar, *Convex analysis and optimization*. Belmont, M.A., USA: Athena Scientific, 2003.
- [31] S. Ye and R. S. Blum, "Some properties of the capacity of MIMO systems with co-channel interference," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, vol. III, Mar. 2005, pp. III–1153 – III–1156.
- [32] S. Kaviani and W. A. Krzymień, "Optimal multiuser zero-forcing with per-antenna power constraints for network MIMO coordination," *EURASIP J. Wireless Commun. Networking*, 2011, Article ID 190461.
- [33] W. L. Stutzman and G. A. Thiele, *Antenna theory and design (2nd edition)*. John Wiley & Sons, 1998.
- [34] H. Huang, O. Alrabadi, J. Daly, D. Samardzija, C. Tran, R. Valenzuela, and S. Walker, "Increasing throughput in cellular networks with higher-order sectorization," in *Proc. Asilomar Conf. on Signals, Systems and computers*, Nov. 2010, pp. 630–635.
- [35] A. Hjörungnes and D. Gesbert, "Complex-valued matrix differentiation: techniques and key results," *IEEE Trans. Inform. Theory*, vol. 55, no. 6, pp. 2740 – 2746, Jun. 2007.